2014

Adaptations Of Bacterial Flagella To Animal And Plant Hosts

Alshaeâ€™ R. Logan

North Carolina Agricultural and Technical State University

Adaptations of Bacterial Flagella to Animal and Plant Hosts

Alshae' R. Logan

North Carolina A&T State University

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Department: Biology

Major: Biology

Major Professor: Dr. Scott H. Harrison

Greensboro, North Carolina

2014

The Graduate School
North Carolina Agricultural and Technical State University
This is to certify that the Master's Thesis of

Alshae' R. Logan

has met the thesis requirements of
North Carolina Agricultural and Technical State University

Greensboro, North Carolina
2014

Approved by:

Scott H. Harrison, PhD Major Professor	Jian Han, PhD Committee Member
Catherine D. White, PhD Committee Member	Mary A. Smith, PhD Committee Member
Mary A. Smith, PhD Department Chair	Sanjiv Sarin, PhD Dean, The Graduate School

Biographical Sketch

Alshae' R. Logan is a native from South Boston, Virginia. She was born February 28, 1988, the oldest of two to the late Allen Raye Logan and Veronica Coleman. She attended the Halifax County School District. She has been in church since an early age. After graduating Halifax County High School, she attended Shaw University where she obtained her Bachelor's Degree of Science in Biology. While getting her bachelors, she was funded by a NIH grant working under the direction of Dr. Mialy Rabe. During the two summers of Shaw, she worked at the Department of Environmental and Natural Resources, working under the branch of Water Quality. A year later, she went on to North Carolina Agriculture & Technical State University to acquire her Master's of Science in Biology. She has been funded by the NSF-BEACON grant and NIH/NIGMS MBRS-RISE under the direction of Dr. Scott Harrison. She has presented at several conferences and seminars delivering her research experiences.

She is a mother of one to the precious Denae Makayla Logan.

Dedication

This thesis is dedicated to my daughter, Denae Makayla Logan, who holds a special place in my heart forever.

Acknowledgments

I would like to thank my thesis advisor Dr. Scott Harrison, for his guidance and encouragement through my years spent at NC A&T SU. He has been a brilliant and wonderful mentor, who challenged me to think critically and put forth my best effort in research. I am more than grateful for his feedback and support for my future endeavors. I am also grateful for the support and advice from members of my thesis committee, Drs. Jian Han, Catherine White and Mary Smith. They have helped me along the way with my experiments and gave me great input on my research. Additionally, I would like to thank NC A&T SU and the Biology Department for financial support through the NSF BEACON training grant and last but not least the NIH/NIGMS MBRS- RISE program.

I am forever thankful for my loving fiancé Danny Jackson who always cared about my education and well-being. He has been there through the rough and good times; I can truly say he is here for the long haul. I am also thankful for my best friend Sabena Conley, who was there to help me with my lessons and laboratory math. I would like to thank my grandmother, Annie Hicks who helped me with my daughter so that I could excel in my studies. She has always helped me financial and mentally. There's nothing like a loving grandmother. I also would like to thank my mother, Veronica Coleman who has always been there from day one. She inspired me to be better than she was and I am grateful for her words of encouragement. She always told me to set my sights higher.

This material is based in part upon work supported by the National Science Foundation under Cooperative Agreement No. DBI-0939454. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Table of Contents

List of Figures

List of Tables

Abstract

Flagella have been studied for more than a hundred years, yet there remains a simple and powerful question about how to further make use of the recent wave of genomic data to find differences between flagellar genes across diverse bacterial strains. With the comparative power of bacterial genomic data, we propose that flagellar gene sequences may be used as genetic markers that indicate evolutionary adaptations of different bacteria for different hosts. We specifically proposed to identify and utilize genetic sequence differences that may diagnose for the presence of bacterial pathogens found to infect different hosts, specifically animal versus plant hosts. Four flagella genes were chosen for this study due to the fact that they are consistently found across diverse types of bacteria having both flagellar phenotypes and fully sequenced genomes. *fliC* and *fliD* express proteins that are located on the outside of the cell and are potentially antigenic in animal hosts. *fliJ* and *flgG* are expressed as intracellular protein products. We collected sequences for these flagellar genes from 18 animal host bacteria and 18 plant host bacteria and analyzed for homologous regions and dN/dS measures of selective pressure. As expected, *fliC* had the highest amount of diversification for animal host bacteria versus plant host bacteria. We developed a quantitative trait loci approach for identifying differentiating regions within the *fliC* gene - a 51 nucleotide region encoding for some of the conserved N-terminal domain of flagellin (*FliC*) and a 21 nucleotide region encoding for the more variable middle domain of flagellin. We then developed and tested an algorithm with these regions against our database of 36 bacterial strains, and the performance characteristics of the algorithm were best for inferring the presence of plant host bacteria compared to animal host bacteria. We implemented an extraction and PCR amplification protocol to screen for bacterial *fliC* DNA from agricultural produce and, for multiple replicates from a single specimen of

lettuce, found sequence patterns corresponding to both the N-terminal domain and middle domain that were consistent with plant host bacteria. We anticipate that further development of the algorithm and the usage of next generation methods of sequencing will help enhance this overall workflow for more extensive contexts of usage.

CHAPTER 1

Introduction

1.1 Flagella

The flagellum is a virulence factor for bacterial pathogens in both animals and plants (Finlay & Falkow, 1997). Flagella genes assist with the export process, motility and the chemotaxis stage. Flagella have been studied for more than a hundred years, yet, there has not been a comprehensive study of evolutionary pressures on flagellar genes between animal and plant host bacteria based on the recent wave of genomic data. A primary factor of comparison to the evolutionary pressures on these four flagella genes in animal versus plant host contexts is the presence of an adaptive immune system in animals.

Four genes of the flagellar apparatus were chosen for this study due to the fact they are well studied in the literature and common to a wide variety of bacteria. For the purpose of comparison, two encode for proteins for the extracellular portion and the other two for the intracellular segment. Some of the extracellular gene products are expected to have interaction and selective pressure in response to host factors such as the antibodies of animal immune systems.

1.2 Flagellin

We mainly focused our study on *fliC*, the gene that encodes for flagellin. Flagellin is a major component of the bacterial flagellum. There are 10,000 to 40,000 molecules of flagellin within the filament of a flagellum. Flagellin is self-assembled at the distal tip after going through the hollow core of the flagellum (Felix & Boller, 1999). The conserved N and C terminal ends of flagellin determine the export and self-assembly. Flagellin is essential for the phenotype of cellular attachment on animal cells where *fliC* attaches to Muc1 mucin. As the extracellular

portion of the flagella interacts with the host cell boundary, the flagella release “effector” proteins into the host cell and foster invasion through some cross-interaction with bacterial type III secretion systems, which is poorly understood (Abby & Rocha, 2012; Alfano & Collmer, 2004). Attachment to plant cells does not rely on a specific ligand receptor mechanism, but rather upon the complex interactions among the diverse materials of bacteria and plant cells (Wilson & Beveridge, 1993; Lillehoj et al., 2002).

1.3 Objectives and Impact

The objective of the current research was to determine evolutionary differences between animal and plant bacteria as they relates to flagellar genes. Based upon the adaptive immune system found in animals and differences in cell attachment mechanisms, we hypothesized that flagella genes from animal host pathogens such as *Pseudomonas aeruginosa* will differ from the plant host pathogens such as *Pseudomonas syringae*. Diversifying selection of extracellular flagella genes that alters amino acid composition may be expected for bacterial adaption to the animal immune system which could otherwise recognize and destroy an unchanging antigenic surface on the bacteiral cell. We expect however some balancing mode of selection to some of these diversifying changes. Flagella are highly interconnected systems, and we postulate that there would a high level of negative (purifying) selective pressure in order to maintain functionality of the interdependencies of quaternary structure and dynamics for the dozens of different types of proteins of a flagellum. We investigated this by first analyzing the flagella genes and proteins of diverse bacteria, utilizing KEGG and IMG databases. For the validation of the genotypic measures which may differentiate bacterial flagella sequences based on evolutionary pressure for animal versus plant host association, we developed initial protocols and workflows for DNA sequencing and phenotypic attachment that may be applied to bacteria

found in nature and the human environment. Our strategy for differentiation may lead to a powerful scenario of application based on how bacteria that cause disease in plants do not cause disease in animals, and vice versa. As those categories are effectively contrasted, this may lead to a generic and simplified screening method in comparison to methods that are specific to narrow ranges of pathogens. This would however rely upon an effective approach for extracting DNA from diverse bacteria potentially found on eukaryotic tissue. These approaches may range from isolation and growth of bacterial cells in culture media to a wholesale metagenomic capture of all DNA from any eukaryotic specimen followed by digital subtraction to detect for bacterial DNA (Duncan et al., 2009). In this study, as an initial step of validation for future directions, our approach was for PCR-based targeted sequencing of the bacterial *fliC* gene from a bulk DNA extraction conducted upon lettuce tissue.

CHAPTER 2

Literature Review

2.1 Bacterial Diversity and Structure

Bacteria are a domain of unicellular microorganisms with a variety of cellular morphologies with shapes ranging from spirals, rods and spheres. In addition, bacterial cells may have external structures such as fimbriae, pili and flagella. A major distinction across this domain has historically been for gram-positive or gram-negative status which generally corresponds to having one or two phospholipid bilayer membranes for their cell boundaries respectively (Gupta, 2011). Gram-negative bacteria are very unique in having two membranes for their cell boundary – an outer membrane and plasma membrane. The volume between the outer and inner membrane has the periplasmic space and a thin peptidoglycan layer. Gram-negative bacteria have porins which are transmembrane proteins that allow passage for certain molecules across the outer membrane (Schirmer, 1998). A chemical signature of gram-negative bacteria is the presence of lipopolysaccharide as a component of their outer membrane. Examples of gram-negative pathogens that infect the human population include strains of *Escherichia coli*, *Salmonella*, *Shigella*, *Stenotrophomonas*, *Pseudomonas* and *Vibrio* (Brooke, 2012; Hunt et al., 2011; Shute, 2013; Haley et al., 2013; McClaine & Ford, 2002; Abby & Rocha, 2012).

For many gram-negative bacteria, their two membranes provide for enhanced flux control of the import and export of small molecules, enabling sophisticated systems of antibiotic resistance, endosymbiosis, and secretory transport that go beyond what is found for gram-positive bacteria (Gupta, 2011; Gerlach & Hensel, 2007). In addition, gram-negative pathogens are more likely to have pili and fimbriae (Proft & Baker, 2009; Winn, 2006). Flagella are

widespread across both gram-positive and gram-negative diversity and this is attributed to their early appearance in the evolution of bacteria several billion years ago (Cavalier-Smith, 2002). These virulence factors of gram-negative bacteria are critical to their pathogenesis in a wide range of host symbioses including gastrointestinal disease.

The effect on eukaryotic hosts of gram-negative pathogens varies from plants to animals. Along with other exterior factors, the lipopolysaccharide (LPS) layer of the gram-negative bacteria outer membrane triggers an innate immune response in animals through binding to TLR-4 receptors of animal cells, ultimately initiating an inflammatory response. The animal host defense system may then generate a secondary adaptive immune response mediated by white blood cells that can acquire the ability to recognize a wider range of foreign material (Aderem & Ulevitch, 2000). For bacterial adaptation to this powerful animal immune system, it has been suggested that there is diversifying selection on genes encoding for extracellular portions of structures such as the outer portions of bacterial flagella (Lynch, 2012). This variation allows the bacteria to evade recognition. This strategy may not be necessary for plant host-associated bacteria. Plants only have a primary (non-adaptive) detection and response system that will enact defense by excreting exopolysaccharides to induce water soaking and wilting of the plant by blocking the xylem (Boch & Bonas, 2001). In order to arrive at a focused comparison of this selective effect on animal versus plant host-associated bacteria, it would be ideal to utilize those gram-negative taxa for which there are multiple strains of both plant and animal pathogens, and for which significant background knowledge exists for the genomes and phenotypes of these strains. One such opportunity for comparison among related gram-negative bacteria that have known host associations and fully sequenced genomes would be a comparison between *Pseudomonas aeruginosa* that infects animals and *Pseudomonas syringae* that infects plants.

2.2 The Genus of Pseudomonas

Pseudomonas is a gram-negative genus of the family *Pseudomonadaceae* of the class *Gammaproteobacteria*. The name *Pseudomonas* was proposed by Walter Migula in 1894 and soon thereafter was adopted in the 1900s literature (List of Prokaryotic names with standing in nomenclature. Retrieved September 5,2013, from http://www.bacterio.net/). Overall, *Pseudomonas* is a gram-negative, aerobic, rod-shaped (bacillus-shaped) bacterial taxon with polar flagella. There are currently 206 species that have been formally described in the literature (List of Prokaryotic names with standing in nomenclature. Retrieved September 5,2013, from http://www.bacterio.net/). The different species of this genus have a wide range of metabolic diversity and are able to colonize a wide range of habitats including terrestrial, aquatic, freshwater and clinical habitats (Hirano & Upper, 2000; Romling et al., 1994). Some species such as *Pseudomonas aeruginosa* affect the animal population with impacts on human and livestock health while others, most notably *Pseudomonas syringae*, affect plants with impacts on agriculture. The evolution of different varieties of *Pseudomonas* has only begun to be interpreted (Baltrus et al., 2011), and opportunities exist for comparing different genomic sequences and investigating their evolutionary divergence as a function of different environments and host associations.

2.3 *Pseudomonas syringae*

Pseudomonas syringae causes frost damage across many different crop species. *P. syringae* genomes are approximately 6 million base pairs (bp) in size. Distinct strains of *P. syringae* only exhibit pathogenic potential on a single or small range of plant species, but for other plant species will fail to initiate disease. For pathogenesis, *P. syringae* strains transfer a large number of different proteins into plants through the bacterial type III secretion system

(T3SS). *P. syringae* has many polar flagella that are used for motility. Three strains of *P. syringae* have been identified and fully sequenced. These three strains have different effects on host plants. *P. syringae* pv. *tomato* DC3000 causes bacterial specks on tomatoes and *Arabidopsis* (Wilson et al., 2002; Whalen et al., 1991). *P. syringae* pv. *syringae* B728a causes brown spots on beans (Willis et al., 1990). *P. syringae* pv. *syringae* 1448A causes halo blight on beans (Taylor et al., 1996). The strain that has been most intensively studied and, advantageously for experimental study, grows on the model plant organism *Arabidopsis* is *Pseudomonas syringae* pv. *tomato* DC3000. (Fouts et al., 2001).

The Hrp-dependent pili protein that affects *Pseudomonas syringae* pv. *tomato* (Pst) DC3000 is HrpA. The Hrp pilus is suggested to be involved in type III secretion of HrpN and DspA/E. Bacteria having knockout mutations in the gene for this protein, *hrpA*, do not cause disease in host plants and do not elicit the hypersensitive response in non-host plants. Researchers have shown through genetic analysis that the *hrpA* gene is required directly or indirectly for the secretion of effector proteins. The *hrpA* gene of Pst DC3000 is necessary for the full expression of *hrp* and *avr* genes encoding for Hrp and Avr proteins respectively (Hu et al., 2001).

2.4 *Pseudomonas aeruginosa*

Pseudomonas aeruginosa is an opportunistic pathogen of animals. For example, it causes chronic and sometimes lethal infections of the lungs of patients with cystic fibrosis (CF) (Lillehoj et al., 2002). *P. aeruginosa* is responsible for a variety of diseases: 16% of nosocomial pneumonia infections, 12% of nosocomial urinary tract infections, 8% of surgical wound infections and 10% of bloodstream infections. *P. aeruginosa* causes death rates reaching 30% with pneumonia patients, and it is among the most common and lethal pathogens for ventilator-

associated pneumonia in intubated patients causing death rates reaching 38%. Within intensive care units, *P. aeruginosa* has shown to reach high death rates approximately 60%. For persons with AIDS, *P. aeruginosa* has been found to have a 50% death rate (Deldan et al., 1998). It can be found in soil, water, skin flora, and most human-made environments throughout different countries all over the world. *P. aeruginosa* infections within animals feed off of a variety of organic material (Lillehoj et al., 2002). *P. aeruginosa* has a versatile set of nutrient uptake pathways and response behaviors including for carbon and nitrogen that modulate in vitro biofilm development and surface motility. Carbon sources have two effects on virulence phenotypes within *P. aeruginosa*: toxin production and formation of antibiotic-resistant biofilms (Palmer et al., 2007). Generalized inflammation and sepsis are symptoms of the infection. Colonization occurs in the lungs, the urinary tract, and kidneys. These infections can be fatal to immunocompromised individuals (Lillehoj et al., 2002).

Pseudomonas aeruginosa has nine fully sequenced genomes that range in size from 6.2 to 6.9 million bp (Winsor et al., 2011). In terms of genomic analysis, PA01 is the most well studied strain (Fouts et al., 2001). The PA01 ATCC 15692 strain's genome is 6,264,404 bp and there are 5,571 open reading frames within this strain (Fouts et al., 2001).

Investigations into the microbial ecology and molecular mechanisms of *P. aeruginosa* have identified patterns of gene expression underlying host and environmental adaptations. During the infection state in *P. aeruginosa, lasR* triggers quorum sensing leading to measurable increases of cell density, virulence factor synthesis, and biofilm production. These phenotypic analyses have been confirmed by disrupting the *lasR* mutants (Smith et al., 2006). Other phenotypic measures of *P. aeruginosa* have been for alginate production, under aerobic versus anaerobic conditions on Columbia sheep blood agar plates. For PA01, alginate production and

biofilm formation have been found to increase under anaerobic conditions, and this has been implicated with anaerobic conditions in the mucus of cystic fibrosis patients (Worlitzsch et al., 2002). Flagella play a major role in the virulence of a strain (Finlay & Falkow, 1997). Components of the multiprotein flagellum structure have been evaluated for their role in virulence such as for attachment to mucin 1 host cell receptors. A series of studies compared the wild type *Pseudomonas aeruginosa* PAK to strains devoid of components such as pilin (PAK/NP), flagellin (PAK/*fliC*), and cap protein (PAK/*fliD*) to binding upon human and mouse cells. Although an initial report implicated cap protein (FliD) as the adhesin to mucin 1 (Arora et al., 1998), it was later determined that flagellin protein (FliC) is the adhesin responsible for binding to mucin 1 (Lillehoj et al., 2002).

2.5 Bacterial Flagella

The flagellum is a virulence factor for bacterial pathogens in both animals and plants (Finlay & Falkow, 1997). A bacterial flagellum is a semi-rigid complex structure which has a rotary motor, powered by a transmembrane proton motive force. A bacterial flagellum consists of a long helical filament, hook and basal body that can enable motility for microbial cells (Schuster et al., 1994; Macnab, 2004; Felix et al., 1999 & González-Pedrajo et al., 2002). The proton motive force causes bacterial flagella to rotate at different speeds. This permits some varieties of bacteria to rapidly propel themselves at many micrometers per second. For example, *Bacillus subtilis* was found to move at a swimming speed of 40 micrometers per second (Manson et al., 1977; Khan & Macnab, 1980). An estimated number of 50 proteins make up the complex formation mechanisms and structural components of the flagellar complex (Kutsukake, Ohya & Lino, 1990).

Flagella can serve as catalysts for both the movement and attachment of pathogenic bacteria to various surfaces such as lungs, soil surfaces, meats, and plants. Flagella enable adhesion through physicochemical characteristics of their amino acids including electrical charge and hydrophobicity. Attachment may further involve entrapment, based on first the ability to swim through crevices and then second, perhaps due to how flagella increase the bulk of the cell for it to remain situated in the constrained space of a crevice (Piette & Idziak, 1991). Flagellar attachment has been linked to subsequent events of pathogenesis. In a study done on flagella-induced immunity against experimental cholera disease in adult rabbits, *Vibrio cholera* was found to bind to an epithelial ganglioside receptor. This study suggests that flagella function to transport and attach *V. cholera* to the intestinal mucosa where subsequent events such as enterotoxin delivery may occur (Yancey et al., 1979).

2.6 Flagellar Bacteria in Plants

In rhizosphere bacteria, flagellar chemotaxis is vital for colonizing the host plant in some instances. A study has shown that flagella were essential for colonization of *P. putida* on potato roots where motility is required to reach the root surface (Broek, 1995). The host resistance of a plant species against specific pathogens that invade it is dependent upon a set of important recognition mechanisms for molecules, through the loss or change of the toxin's target or through detoxification, that are generated exclusively by particular strains of pathogens. R genes enable plants to detect avirulence genes, and to initiate signal transduction that helps with defenses (Hammond-Kosack & Jones, 1997; Parniske et al., 1997). Plants do not have an immune system comparable to vertebrate animals, however plants have the ability to sense infection by conserved molecular aspects of microbial pathogens, such as the N-terminal region

of flagellin. Plants respond to the attack with an extensive set of protection responses (Felix & Boller, 1999; Chesnokova et al., 1997; Hatterman & Ries, 1989).

2.7 Flagellar Bacteria in Animals

In animal-associated pathogens, the flagellum is a virulence factor which has been examined in several animal models to determine the efficiency of the motility directed by flagella. Flagella are responsible for adhesion by virtue of physicochemical properties, electrical charge and hydrophobicity (Piette & Idziak, 1991). Flagella provide a means for some bacteria to colonize throughout the respiratory tract, and associated chemotaxis guides the organisms toward preferred substrates. For *P. aeruginosa*, these preferred substrates include amino acids, inorganic phosphate, and other known components of human mucus (Feldman et al., 1998). In terms of host defense response, flagella are adhesins for polymorphonuclear leukocytes (PMNs) and for macrophages which clear micro-organisms from mucosal surfaces (Mahenthiralingam & Speert, 1995). The general process of the role of flagella in adhesion of *P. fluorescens* has been demonstrated on tendon slices. The presence of flagella promotes motility by modifying the chances of bacterium to encounter the favorable location and possibly provides the kinetic energy to initiate adhesion. The flagellum then initiates contact of the bacterial cell to the meat (Piette & Idziak, 1991).

2.8 Motility and Attachment

Motility associates with virulence, biofilm development and invasiveness in *Pseudomonas syringae* and *Pseudomonas aeruginosa* (Romantschuk, 1992; Turnbull & Sanders, 2001). Motile bacteria move to favorable locations through chemotaxis. *Pseudomonas aeruginosa* cells aggregate and form biofilms under conditions that allow for growth, such as availability of oxygen, sugars and amino acids. Flagella have been shown to play a vital role in

the initial event of biofilm development in *P. aeruginosa*. Once bacteria come into contact with the surface, they may attach and form biofilms (O'Toole et al., 2000). In natural environments, non-motile and motile bacteria appear at their favorable locations because they stick to surfaces (Romantschuk, 1992). Motility allows a significant amount of cells to get closer to the surface, however a necessary pause is required near the surface to allow the host cell to attach effectively (McClaine & Ford, 2001).

The initial attachment of cells occurs in two steps, transport to the surface and absorption to the surface. Motile cells swim in all directions and non-motile cells do not. Non-motile cells attach at a much lower rate than active motile cells because motile activity facilitates cells to contact attachment areas more quickly; also active motility increase the probability of reaching a potential binding site rather than relying on being stagnant. Motility increases the chance to position the cell in close enough physical proximity to the host tissue for binding to occur (McClaine & Ford, 2001).

Studies have shown the attachments of bacteria to surfaces are measured by the fluid velocity by comparing motile bacteria to nonmotile bacteria (McClaine & Ford, 2001). Swimming bacteria behave differently than non-swimming bacteria in that motility assists transport through permeable media in dormant and flowing systems and also favors attachment to surfaces in smaller-scale systems under vibrant conditions (Jenneman, McInerney & Knapp, 1985; Reynolds et al., 1989; Witt et al., 1999; Camper et al., 1993; Korber, Lawrence & Caldwell, 1994; Lawrence et al., 1987; Mueller et al., 1992; Mueller, 1996). Measurements have involved the residence time analysis, motile bacteria have time to reach the surface at different flow rates because they have the ability to swim at higher speeds than non-motile cells (McClaine & Ford, 2001).

Attachment in bacterial cells to foliage and to the root surface in many plant hosts is an early vital step. It is a significant aspect in epiphytic colonization; researchers have thought this increased the ability of pathogenic strains to result in plant disease. Surface openings, stomata and wounds in plant tissue are the way for bacteria to enter into the host cell. Attachment to the epidermal surface offers a selective advantage for epiphytic colonizers, although it does not directly induce a plant defense response as vertebrates does (Boch & Bonas, 2001). Specific chemotactic attractants such as galactose, glucose, arabinose, fucose and xylose can stimulate virulent gene expression for bacteria associated with detecting and entering plant wounds. Many substances are in plant root exudates and these include amino acids and simple sugars along with specific nonmetabolizable compounds which are identified as chemoattractants for different plant host bacteria. Bacterial cells within the root surface microenvironment are increased because of motility which facilitates transport from the aqueous bulk phase surrounding the root (Turnbull & Sanders, 2001).

Once bacteria have been embedded in the plant roots or the seed, flagella can act to move bacteria to favorable locations for attachment (Broek et al., 1995). The extracellular portion of flagella interact with the host and then the flagella release “effector” proteins into the host cell and foster invasion through some cross-interaction with bacterial type III secretion systems, which is poorly understood (Abby & Rocha, 2012; Alfano & Collmer, 2004). Researchers have found that motility and chemotaxis initiate the interaction of rhizobia with the plant root surface. This mechanism ensures the effectiveness of inoculation and improves the speed of plant growth (Turnbull et al., 2001). Once on the surface of the root, bacteria can be passed downward by the roots reaching into the soil of the plant. During this stage, flagellar motility and chemotaxis

become a part of the process which then spreads the bacteria in the rhizoplane where root mucilage contains an ongoing water film.

Many plant hosts are infected with bacteria that are motile in the free-living state. Rhizosphere bacteria are flagella-driven, which is essential to their virulence (Felix & Boller, 1999). Flagella of rhizobacteria constitute the mechanism for chemotaxis and motility, which can lead to early contact (Broek et al., 1995). Motility is vital for foliar pathogens reaching internal sites in the leaves of host plants. Some plants have a highly sensitive chemoperception for the N-terminal domain of bacterial flagellin. *A. thaliana* and tomato cells have been found to respond to flagellin, unlike rice cells. Peptides that consist of 15 through 22 amino acids within the N terminus domain "acted as elicitors of defence responses at subnanomolar concentrations" (Felix & Boller, 1999) in tomato cells and other plant species. Peptides that consist of 8 through 11 amino acids act as direct competitive inhibitors of defense responses in tomato cells but do not have an elicitor effect within this domain (Felix & Boller, 1999).

A particular study demonstrated attachment on *P. putida* to wheat roots by comparing motile and non-motile strains of *P. putida* on the same root section (Turnbull et al., 2001). The non-motile strains attached at a lower number suggest that the motile strain was competing for the same binding site. The motile strain had an advantage because of the flagella which makes this strain motile. Turnbull and his research team in 2001 recommend that perhaps motility can increase the number of bacterial cells that are in the root surface microenvironment by means of facilitating transport from the aqueous bulk phase surrounding the root. They further propose that motility can overcome electrostatic repulsion or surface tensions which are physicochemical forces.

The first step of binding of bacterial cells to plant tissue starts off by early reversible and irreversible attachment steps which are specific to the cell interaction. An electrostatic force keeps the cells apart in an aqueous environment possessing a net negative charge. The only way for the bacteria cell to attach to the plant surface is through ionic bridging between nearby positively charged groups on one surface, hydrogen bonding between suitable groups on either surface, or through attractive close range (3-4 A) van der Waals forces and divalent cations, all of which help defeat the repulsive energy obstacle (Romantschuk, 1992).

Receptors are proteins that found inside or on the surface of the cell, transport signals coming from outside of the cell. There have not been a specific plant cell receptor that bacteria bind to; however, the findings from (Gurlitz et al., 1987) used a few monosaccharides for proteolytic or chemical treatments of the carrot cells to see if there was any binding specificity found in the carrot cell. Sugars in the plant cell wall such as arabinose and galactose failed to inhibit binding of the bacteria, and this confirmed the adherence of Agrobacterium tumefaciens to carrot cells involving a receptor site that is different from other bacteria. This was due to the comparison between the embryos and uninduced cells utilizing that demonstrated changes. The carrot cells were not killed by the treatments and they seemed to recover their binding ability after being in the incubation for 3 to 6 hours. A specific receptor has not yet been found. (Gurlitz et al.,1987).

Colonization is the initial step of microbial infection in animals. In animals, there are many ports of entry that can be infected by bacteria which include the digestive tract, urogenital tract, respiratory tract and the conjunctiva. Flagellar-mediated attachment of *P. aeruginosa* cells to mucins is known for being an important factor in colonization of respiratory epithelial cells (Lillehoj et al., 2002).

Bacterial that attaches to different surfaces forms a slime condition, subsequently forming a biofilm (Costerton, Stewart & Greenberg, 1999), for example the mucosa of the respiratory tract and gastrointestinal tract. Flagella act as ligands to receptors on the surface of animal host tissue cells (Arora et al., 1997). Bacteria infecting animals have a different adhesion process than for infecting plants. *Campylobacter jejuni* was compared to observe the attachment to the gastrointestinal tract using a flagellate motile strain and a flagellate non-motile strain, comparing it with a non-flagellate strain. In order for a successful infection to occur, enteropathogenic bacteria have to overcome host defenses. The gastrointestinal tract mucus layer removes unattached micro-organisms, but motile bacteria can navigate the mucus layer and live under the mucus layer within the dormant area near the epithelial cells (Newell, McBride & Dolby 1985).

Flagellar motility is a general virulence factor of *V. cholerae* essential for transport to the intestinal mucosa. A comparison of motile and non-motile strains of *Vibrio cholerae* established that motility was essential for subsequent adhesion to isolated brush borders and surfaces of mucosal and for agglutination of red blood cells. After being transported, flagellar attachment aids the delivery of enterotoxin to eukaryotic host cells (Newell, McBride & Dolby, 1985).

2.9 Flagellar Genes and Proteins

Flagellar genes that make up the flagellum assist in the ATP driven export process, motility and the chemotaxis stage (Kutsukake, Ohya & Lino, 1990; Macnab, 2004). It is thought that there are 40 to 50 genes essential for the flagellar phenotype. These genes are regulated by sigma factors that control the expression of flagella genes along with other regulation factors that assist with these sigma factors (Arora & Ramphal, 1997). Flagellar gene regulation has been found to occur through a regulon of 13 operons in *Salmonella typhimurium* (Kutsukake, Ohya & Lino, 1990; Komeda, 1982; Poggio et al., 2007).

The flagella proteins are grouped into regions I, II and III. In region I, the hook and basal body proteins are located (Macnab, 1992). The basal body consists of four rings (L, P, S and M) which are all located on a central rod (Aizawa et al., 1985; DePamphilis et al., 1971). Within the rod, there are several proteins (FlgB, FlgC, FlgF,and FlgG), which are all related to each other in primary sequence (Homma et al., 1990), the outside ring (FlgH and FlgI) (Jones, Homma & Macnab, 1989), the hook (FlgE) (Homma, DeRosier & Macnab, 1990) and the hook-filament junction (FlgK and FlgL) (Homma, DeRosier & Macnab, 1990; Homma, Kutsukake & Lino, 1984). Region I is helpful to cell-proximal elements of the external flagellar structure. The chemotaxis proteins are within in Region II. The chemotaxis proteins are CheA, CheB, CheR, CheW, CheY, and CheZ. For the receptor proteins tar and tap, tar is found only in *E. coli* and the motility proteins, MotA and MotB are also found in this region. These chemotaxis proteins are utilized for detecting chemical stimuli, the cells have several receptors, which bind ligands and transport this information to the cytoplasm (Macnab, 1992). The master operon protein is also encoded in this region, FlhD and FlhC (Bartlett, Frantz & Matsumura, 1988) along with the protein that is involved in the export apparatus, FlhA (Ibuki et al., 2011; Vogler et al., 1991; Kim, 1989). Region IIIa is dedicated to the flagellar filament, FliC, flagellin (Joys, 1985). FliA is the protein for an **σ** factor that initiates transcription of *fliC* (Ohnishi et al., 1990). Region IIIb comprises the proteins that are involved in the preliminary phases of the assembly of the flagellum (Macnab, 1992).

For this research, we have selected to study four flagella genes, *fliD, fliC, fliJ* and *flgG*. We are choosing to compare these four genes, being that two come from the external portion, *fliD* and *fliC,* whereas for the other two, one is in the cytoplasm, *fliJ*, and *flgG*, is in the basal body of the flagellum.

2.10 FliD

FliD is involved in filament assembly and ranges in *Pseudomonas aeruginosa* and *Pseudomonas syringae*, from 474 to 478 and 483 to 492 amino acids, respectively (Kyoto Encyclopedia of Genes and Genomes. Retrieved January 29, 2014, from http://www.genome.jp/kegg/pathway.html). It acts as a cap protein for the distal, extending end of the flagellar filament which is critical in the assembly of the FliC, flagellin proteins (Macnab, 1992; Tasteyre et al., 2001). FliD is also referred to as hook-associated protein 2 (HAP2) (Tasteyre et al., 2001; Arora et al., 1998). An alignment analyzed the amino acid sequence for FliD in *P. aeruginosa* in comparison to the other FliD proteins, and demonstrates that the organization of this protein is conserved throughout the ORF (Arora et al., 1998). Normally, *fliD* gene has two chaperone proteins, FliS and FliT, that assist with its function but in *P. aeruginosa* it is different (Arora et al., 1998). There was no evidence of FliT and there was a duplication of the gene encoding for the FliS protein. In *P. aeruginosa*, *fliD* exclusively maintains equilibrium of the motility and mucin adhesion. Research shows that the *P. aeruginosa fliD* knockout mutant did not need the help of the two chaperones to fulfill the two functions of motility and adhesion. The promoter of *fliD* binds to the transcriptional regulator, FleQ and the sigma factor RpoN. The transcriptional regulator, FleQ, works together with the sigma factor RpoN regulating motility and mucin adhesion in *P. aeruginosa.* RpoN has two sigma factors, 54 and 70, that are a dual regulation control of *fliD* expression. Sigma factors aid the flagella gene with motility and adherence. Sequence analysis of the operon containing *fliD* has been done and it showed *fliS* and *fliT* share the same operon as *fliD* in *P. aeruginosa* (Arora et al., 1998). FliD is located in region IIIa (Yokoseki et al., 1995). Another study has been done on *S. typhimurium* with this *fliD* gene

to see if the other two genes assisted with the function of *fliD* and they obtained similar results (Homma, DeRosier & Macnab, 1990).

2.11 FliC

FliC is the major structural component of the flagellar filament. PCR was a method that was utilized to investigate the conservation of the *fliC* gene. The *fliC* gene was sequenced in three strains of *Clostridium difficile*: C, D, and X. Results showed that the analysis in *C. difficile* strains has conservation in the N and C termini, but the middle was very diverse (Tasteyre et al., 2000). In *Pseudomonas aeruginosa*, the amino acid lengths range from 387 to 488 whereas in *P. syringae* an amino acid length of 282 has been consistent across all strains (Kyoto Encyclopedia of Genes and Genomes. Retrieved January 29, 2014, from http://www.genome.jp/kegg/pathway.html). This is an important gene because it exhibits the behavior of mediating flagella endocytosis which assists with bacterial "effector" proteins entering into the host cell. Studies have shown the knockout gene *fliC* of *B. pseudomallei* did not demonstrate flagellum-mediated endocytosis. It was timed in coculture and it confirmed that only an intact flagellar machinery supported *B. pseudomallei* access into *A. astronyxis* (Inglis & Chang, 2003). Lillehoj et al (2002) showed that mucin 1 on the epithelial tissue surface is an adhesion location for *P. aeruginosa*. This study demonstrated that flagellin is an adhesion of *P. aeruginosa* accountable for binding to mucin 1; *fliC* mutants are non-motile and do not adhere to mucin receptors of host cells (Lillehoj et al., 2002).

Researchers have observed the NHO1 gene in *Arabidopsis* is necessary for resistance to numerous strains in non-host *P. syringae.* Results demonstrated that flagellin is the primary Pathogen Associated Molecular Pattern (PAMP) in the *Ptab* strain of *P. syringae* responsible for NHO1 induction; the *fliC* mutant strain does not exhibit NHO1 induction. The *Ptab* strain shows

partial virulence on *Arabidopsis* plants that is directly infiltrated into leaves because it lacks the *fliC* gene and also on tomato plants; however the *fliC* mutant is not fully pathogenic on *Arabidopsis* (Li & Zhou, 2005).

2.12 FliJ

The chaperone gene *FliJ* is one of three soluble proteins (FliH, FliI and FliJ) of the export apparatus. The export apparatus has an additional six other membrane proteins, FlhA, FlhB, FliO, FliP, FliQ and FliR (Ibuki et al., 2011). From various studies of *Salmonella*, the length of FliJ is 147 amino acids with a physical protein size of 17 kDa (Vogler et al., 1991; Fraser et al., 2003). FliJ directly interacts with FliH and FliI and with soluble domains of FlhA and FlhB (Minamino et al., 2000). FliI function is to drive the export process by energy from ATP. FliH functions as a negative regulator of FliI (Fraser et al., 2003). *FliH, fliI* and *fliJ* genes are related through evolution to major components of F1-ATPase. FliH, FliI, and FliJ are not absolutely essential for flagellar protein export – flagellar assembly can still occur in their absence (Ibuki et al., 2012). It has been found from a negative dominant study with glutathione *S*-transferase (GST) affinity chromatography and FliJ mutations that FliJ(F72A) and FliJ(L76A) reduce proton motor force-driven export by preventing the wild-type activity of FliJ to connect together two intraprotein domains of FlhA (Ibuki et al., 2012). The inhibition of GST-FliJ on export has been found to be reduced by the mutations, and it has been overall proposed that the FlhA binding surface of FliJ relies upon Gln38, Leu42, Tyr45, Tyr49, Phe72, Leu76, Ala79, and His83 (Ibuki et al., 2012).

FliJ chaperone protein binds to both filament and rod/hook substrates and filament proteins. Filament substrates are divided into two classes: the rod/hook export class and the filament export class (Minamino et al., 2000). The six membrane proteins are found to be

essential for all flagellar protein substrates, which translocates across the plane of the cytoplasma membrane (Minamino & Macnab, 1999). The filament proteins are FlgK (first-hook filament junction protein), FlgL (second-hook-filament junction protein, and FliD (filament capping protein). Various flagellar proteins in *Salmonella* such as FliS, FlgN, and FliT have the same filament substrate chaperoning mechanism as FliJ does. In order for *fliJ* chaperone gene exhibit its complete function, the N-terminal has to have its first 73 amino acids within FliJ, otherwise swarming was reduced. *FliJ* has a unique structure within its gene; it has a sequence with a high probability of alpha helical coiled-coil near the N terminus, which is necessary for the function of this gene. This study found that when FliJ-N73 is overproduced, it stimulates the export of FliC (flagellin) along with FlgD (hook-capping protein). Yet, deletions of residues 13 to 24 reduced motility even in the presence of FliJ overexpression. It is hypothesized that the essential role of FliJ is to prevent aggregation of export substrates in the cytoplasm (Minamino et al., 2000). Literature shows that half of the C-terminal of FliJ associates with the N-terminal of FliH. When they act together differently under noninducing conditions in wild-type *Salmonella,* it inhibits motility. In this study, it demonstrates that the *fliJ* chaperone gene (74-147) truncate binds to FliH in Ni affinity chromatography assays and affinity blotting; the FliJ (1-73) truncate does not bind. None of the 10 amino acid deletions within the C-terminus of FliJ eliminated the FliH binding site. This therefore specifies that the association border may be broad, although those deletions including residues of 101 to 110 had the strongest effect implying that the C-terminal displays the importance of the FliH binding site (Fraser et al., 2003).

2.13 FlgG

FlgG is embedded in the basal body of the flagella as part of the central rod structure that is surrounded by the M, S, P and L rings. Approximately 26 subunits of FlgG comprise the distal

component of the rod that transmits torque to the outer structure of the flagella (González-Pedrajo et al., 2002). The protein length and expression of FlgG can vary. The protein length of FlgG across 411 bacterial strains ranges from 98 to 423 amino acids with a mean length of 262.3 ± 1.87 (95% confidence limits) amino acids (Annotation Search Report on FlgG, Comprehensive Microbial Resource, Retrieved January 18, 2014, from http://cmr.jcvi.org/). Although there is some variation in the composition of operons, the genes for *flgG* and other flagellar proteins are generally found in a tightly organized system of cotranscription across other bacterial species such as *Escherichia coli* and *Salmonella enterica* serovar Typhimurium (Kutsukake, Ohya, & Iino, 1990).

2.14 Summary of Genotypic and Phenotypic Analyses in Flagella

There are opportunities for further investigations of the differences between flagella genes in plant and animal bacteria. Attachment to host cells is an essential phenotypic outcome of bacterial flagella, and its mechanism and rate of incidence may vary. Upon infection, flagella are recognizable by host defense systems and there may be varying levels of diversifying versus purifying selection for some of the flagellar genes. This may enable better understanding of how specific flagellar gene sequences may indicate pathogenicity for potential eukaryotic hosts.

The *fliC* gene has been previously proposed to be a genetic markersfor epidemiological and phylogenetic analysis (Tasteyre et al, 2000). We are going to extend upon this to analyze four flagella genes of bacteria, *fliC, fliD, fliJ* and *flgG*, and evaluate for their evolutionary differences with animal versus plant host associations. Following a bioinformatics analysis, we are going to then develop a PCR-based protocol to be used in a diagnostic workflow to measure the contamination of plant material based on the patterns of identified flagellar gene sequences.

CHAPTER 3

Methodology

3.1 Comparative Sets of Bacteria Strains

There are 18 strains animal host bacteria and 18 plant host bacteria strains that were analyzed for this study, identified from the IMG database http://img.jgi.doe.gov/, January 23, 2013 with cross-referencing to the KEGG database http://www.genome.jp/kegg/pathway.html September 5, 2013 and selected based upon gram negative status and unambiguous association with disease. The animal pathogen strains were: *Edwardsiella tarda* EIB202, *Edwardsiella tarda* FL6-60, *Escherichia coli* BL21-Gold(DE3)pLysS AG, *Pseudomonas aeruginosa* PA7, *Pseudomonas aeruginosa* PA01, *Pseudomonas aeruginosa* UCBPP-PA14, *Salmonella enterica* subsp. *enterica* serovar Choleraesuis, *Shigella flexneri* 301 (serotype 2a), *Shigella flexneri* 2002017 (serotype Fxv), *Vibrio cholerae* IEC224, *Yersinia enterocolitica* subsp. *palearctica* 105.5R(r), *Yersinia pestis* A1122, *Yersinia enterocolitica* subsp. *palearctica* Y11, *Yersinia pestis* Angola, *Yersinia pestis* Antiqua (biovar Antiqua), *Escherichia coli* APEC O78: APECO78_13405, *Salmonella enterica* subsp. *enterica* serovar Typhimurium SL1344: SL1344_1889, and Vibrio sp. EJY3: VEJY3_03600. The 18 plant pathogen strains were: *Erwinia amylovora* ATCC 49946, *Erwinia pyrifoliae* DSM 12163, *Erwinia pyrifoliae* Ep1/96, *Pseudomonas syringae* pv. syringae B728a, *Pseudomonas syringae* pv. phaseolicola 1448A, *Pseudomonas syringae* pv. tomato DC3000, *Ralstonia solanacearum* GMI1000, *Ralstonia solanacearum* CMR15, *Xanthomonas oryzae* pv. oryzae MAFF311018, *Xanthomonas axonopodis* pv. citri 306, *Xanthomonas albilineans*, *Xanthomonas oryzae* pv. *oryzicola*, *Xanthomonas campestris* pv. *campestris* ATCC 33913, *Xanthomonas campestris* pv. *campestris* B100, *Xanthomonas axonopodis* pv. *citrumelo* F1, *Xanthomonas oryzae* pv. *oryzae* PXO99A,

Xanthomonas campestris pv. *campestris* 8004 and *Xanthomonas oryzae* pv. *oryzae* KACC 10331.

3.2 Flagella Genes

There are 4 flagella genes that were analyzed for this study, *fliC, fliJ, flgG* and *fliD* across the 18 animal and 18 plant-host bacterial strains. The genes were chosen from the KEGG interface (http://www.genome.jp/kegg/ retrieved on September 05, 2013) obtaining the nucleotide sequences and the protein sequences of flagellar genes for each strain of bacteria.

3.3 Sequence Alignment Analysis

We used EMBOSS Transeq (http://www.ebi.ac.uk/Tools/st/emboss_transeq/) (Rice, Longden & Bleasby, 2000) to convert DNA sequence of each strain to amino acid sequence. We then made the amino acid alignment using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) (McWilliam et al., 2013). We utilized PAL2NAL (http://www.bork.embl.de/pal2nal/) to generate a codon-based alignment from the amino acid alignment and the FASTA file of DNA sequences (Suyama, Torrents & Bork, 2006). After those crucial steps, we visualized the codon-based alignments with Jalview2 (Waterhouse et al., 2009). Lastly, the codon alignment text file was edited to remove terminating stop codons and was analyzed with HyPhy utilizing the FEL mathematical analysis (http://www.datamonkey.org/) to measure dN/dS (Delport et al., 2010). All four flagella genes codon alignments were visualized in Jalview 2 from the 18 animal and plant bacterial strains. MAFFT was used to insert test case sequences into pre-established background reference set codon-based alignments (Katoh & Standley, 2013). Quantitative trait loci analysis was conducted as described in 3.4.

3.4 Quantitative Loci Trait Algorithm

Quantitative Trait Loci (QTL) is an approach to infer likelihood for sequence variation to associate with phenotypic traits (Kearsey, 1998). We designed a QTL approach to model estimations of conserved patches that are predictive for animal versus plant host association. The first stage of our QTL algorithm is to calculate odds ratios across each and every nucleotide position. Tables 1-4 present a simplified example of how our QTL algorithm works for calculating odds ratios of sequence variation at each nucleotide position for host association. Nucleotide position #4 (n4) represents the strongest contrast of host association (odds ratio is 16:1), and nucleotide positions #2 and #3 (n2 and n3) represent the absence of any contrast between sequence differences and host association (odds ratio is 1:1). Nucleotide position #3 has an odds ratio of 9:1). Two of the rows of Table 1 (boldface row names) present the consensus sequences corresponding to animal versus plant host associations for this example. The second stage of our QTL algorithm was to measure the strength by which different test cases of sequence regions would align and match to the consensus sequences for the alignment regions. This matching score is calculated as the average of the added or subtracted logarithms of odds ratios for which there is identity with the respective animal (+) or plant (-) consensus nucleotide at a given position.

Table 1

Simplified example of contrasted alignment regions of four nucleotides across eight sequences from four animal host bacteria (AHB) and four plant host bacteria (PHB), along with consensus sequences and three test cases for matching score calculation

	n1	**n2**	**n3**	**n4**
AHB #1 (ACTA)	A	C	T	A
AHB #2 (GCTA)	G	C	T	A
AHB #3 (GCTA)	G	C	T	A
AHB #4 (GCTA)	G	C	T	A
PHB #1 (ACTC)	A	C	T	C
PHB #2 (ACTC)	A	C	T	C
PHB #3 (GCTC)	G	C	T	C
PHB #4 (ACTC)	A	C	T	C
AHB consensus (13)	**G**	**C**	**T**	**A**
PHB consensus (15)	**A**	**C**	**T**	**C**
Test case A	G	C	T	A
Test case B	G	C	T	C
Test case C	A	C	T	C

Table 2

Odds ratio for nucleotide position #1 (n1) on simplified example

	Animal host bacteria strains	**Plant host bacteria strains**
Match AHB consensus	3	1
Match PHB consensus	1	3

Table 3

Odds ratio for nucleotide positions #2 and #3 (n2 and n3) on simplified example

	Plant host bacteria strains	**Animal host bacteria strains**
Match AHB consensus	4	4
Match PHB consensus	4	4

Table 4

Odds ratio for nucleotide positions #4 (n4) on simplified example

	Animal host bacteria strains	**Plant host bacteria strains**
Match AHB consensus	4	0 (1)
Match PHB consensus	0 (1)	4

The matching scores for test cases A, B, and C from Table 1 are calculated as follows:

Test case A (perfect match with animal consensus): $(\log(9) + \log(16)) \ / \ 4 = 1.24$

Test case B (variable match): $(\log(9) - \log(16)) \ / \ 4 = -0.575$

Test case C (perfect match with plant consensus): $(-\log(9) - \log(16)) \ / \ 4 = -1.24$

In summary, the matching scores quantify animal host association likelihood as a positive value and plant host association likelihood as a negative value, where the strength of association corresponds to the magnitude of the averaged matching scores.

The third stage of the QTL algorithm is to identify those regions of the test cases which have the highest magnitudes of matching scores for both animal and plant host association likelihoods. These regions were then retrospectively tested against the original sets of animal and plant host bacterial strains, with matching evaluated with BLASTN (window size of 7, *E* values < 0.05), and performance tabulated as sensitivity, specificity and predictive value.

3.5 Primers

Jalview2 was used to visually identify homologous regions at the ends of the aligned sequences that would be generic for amplifying fliC sequences from the range of 36 bacteria. Primers were: 5' - GAIAIACIGTCGTTIGCGTT - 3' (forward) and 5' - TGGCICAIGCIAACCAG - 3' (reverse).

3.6 Reviving Pseudomonas aeruginosa and Pseudomonas syringae

Heat the tip of the outer vial in a flame, then squirt few drops of water on the hot tip. Strike the tip with a pencil to remove the tip, then remove insulation and inner vial with forceps. Aspectically add 0.5ml of Luria Broth to the freeze-dried material with a sterile pipette and mix well; transfer the total mixture to a 5ml falcon tube. Lastly, incubate *aeruginosa* at 34C and *syringae* at 28C.

3.7 DNA Extraction from Bacterial Cells

Pelleted bacterial cells from 100 µl of quantified *P. aeruginosa* and *P. syringae* were used for DNA extractions. A 300- µl mixture containing Tissue and Cell Lysis solution and 1 µl proteinase K, was added to each sample and the samples were mixed thoroughly by vortex the

tubes for 10 seconds. The samples were incubated at 65°C for 15 minutes; vortex every 5 minutes. The samples were placed on ice for 5 minutes and then proceed with 150 µl protein precipitation reagent and then centrifugation at 4°C for 10 minutes at ≥10,000 x g in a microcentrifuge. The supernantant was transferred to a clean microcentrifuge and discard the pellet. For the precipitation of DNA, 500 µl isopropanol was added to the sample supernatants and were inverted 30 times. The samples were centrifuged at 10,000 g for 10 min. The isopropanol was poured off carefully, then the DNA pellets were washed twice with 75% ethanol and resuspened in 35 µl of DNA Free H2O **.**

3.8 DNA Extraction from Agricultural Produce

2 mg of lettuce were processed from head and bag lettuce samples coming from two different grocery stores. A 300- µl mixture containing Tissue and Cell Lysis solution and 1 µl proteinase K, was added to each sample and mix thoroughly. We homogenize fresh tissue by freeze the tissue using liquid nitrogen. The samples were incubated at 65°C for 15 minutes; vortex every 5 minutes. The samples were placed on ice for 5 minutes and then proceed with 150 µl protein precipitation reagent and then centrifugation at 4°C for 10 minutes at ≥10,000 x g in a microcentrifuge. The supernatant was transferred to a clean microcentrifuge and discard the pellet. For the precipitation of DNA, 500 µl isopropanol was added to the sample supernatants and were inverted 30 times. The samples were centrifuged at 10,000 g for 10 min. The isopropanol was poured off carefully, then the DNA pellets were washed twice with 75% ethanol and resuspened in 35 µl of Tris-EDTA buffer**.**

3.9 Series of Serial Dilution Inoculation

We obtained the revived *P.aeruginosa* and *P.syringae* grown in turbid culture for 48h+ and conducted serial dilutions and plating for 1:10^4 to 1:10^9 with incubation for 72 hours at 28C and 37C respectively.

3.10 Induction of Bacteria into Combined Lettuce

We combined lettuce specimens and inoculated it with *P.syringae* and *P.aeruginosa* using two different mortar and pestals, with a inoculation time of 90 minutes. Then we extracted DNA.

3.11 Mouse Cell Culture

Cell came from ATCC, Hepa-1c1c7 ATCC CRL-2026. The cells were grown in Alpha minimum essential medium without nucleosides, 90%; fetal bovine serum, 10%.

3.12 Optimization-of-Attachment and Cell Culture Wash

The mouse cell line C1C17 was incubated with *P. aeruginosa* and *P. syringae,* separately. The lettuce tissue was also incubated with both bacteria seperately. They were incubated for 30 minutes on a depressed microscope slide. There was a volume of 100uL of bacteria and 100uL of C1C17. The lettuce tissue was 1gram and 100uL of bacteria. The unattached bacterial cells and lettuce tissue and mouse cells were pressed down with a coverslip and poured into an empty beaker. We then took the droplet containing both precipitated eukaryotic cells and bacterial cells were pressed down with a coverslip, then the attached cells and tissue were rinsed three times into an empty petri dish. We streaked duplicate plates for each bacterium containing the mouse cells and the lettuce tissue. The plates were incubated for 48 hours and the colony forming unit counts of bacterial suspensions were 471 cells per mL for *Pseudomonas syringae* and 72 cells per mL for *Pseudomonas aeruginosa.*

CHAPTER 4

Results

4.1 Differences across Flagellar Genes for Animal and Plant Host Bacteria

There was a difference in conservation between the flagella genes (Tables 1 and 2). The *fliC* gene varied the most in length (>20% across 95% confidence interval for plant host bacterial *fliC* genes and >25% across 95% confidence interval for animal host bacterial *fliC* genes). Based on both alignment of sequence homologies and ratios of positive versus negative dN/dS positions, the interior genes *fliJ* and *flgG,* and the exterior gene *fliD* (alignment was essentially uninterrupted), for both animal host and plant host bacteria were mostly conserved. *FliC* was conserved for plant host bacteria but diversified for animal host bacteria (Figure 1). In the *P. syringae, fliC* from nearly all the strains were conserved. Figure 1 shows the locations of the positive selection instances for nucleotide positions where dN/dS > 1 and the negative selection instances for nucleotide positions where dN/dS < 1. The variation in length is also evident, especially from the presence or absence of different subsequences in the middle region. This patchiness of different subsequences in the middle region was entirely absent from sequence alignments for *fliD*, *fliJ* and *flgG*.

Table 5

Profiles of flagellar genes for 18 animal-host bacteria

Gene	Length (95% CI)	Positive (dN/dS > 1)	Negative (dN/dS < 1)
fliC	1315 (1153, 1476)	5	193
fliD	1546 (1437, 1656)	1	236
fliJ	442 (426, 459)	2	76
flgG	800 (799, 801)	0	170

Table 6

Profiles of flagellar genes for 18 plant-host bacteria

Gene	Length (95% CI)	Positive (dN/dS > 1)	Negative (dN/dS < 1)
fliC	1110 (1004,1216)	2	133
fliD	1454 (1341,1568)	0	7
fliJ	456 (453,459)	2	64
flgG	786 (785, 787)	0	130

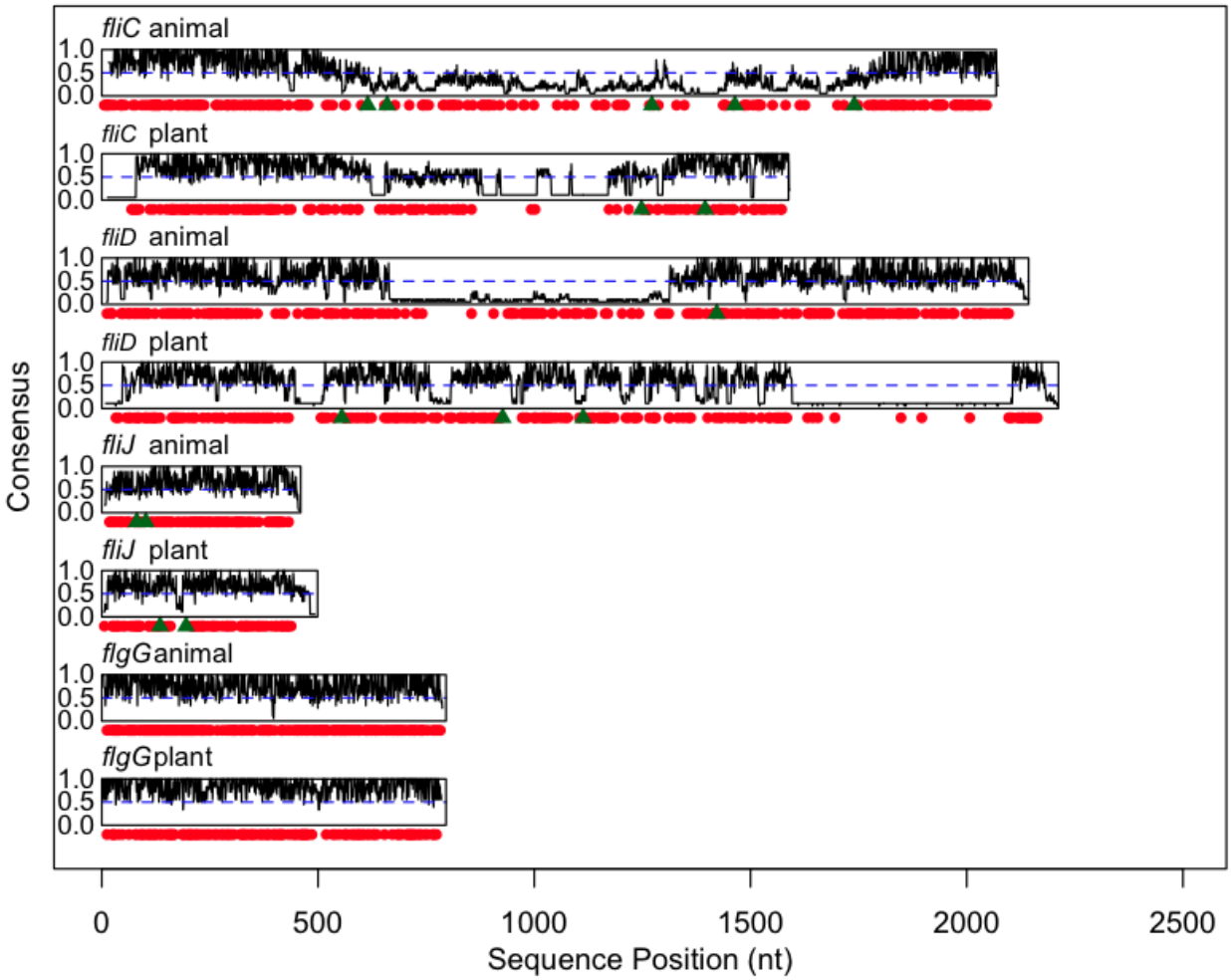

Figure 1. Four flagella genes investigated in animals and plant hosts: two from the extracellular portion, *fliC*, *fliD*; and two from the intracellular portion, *flgG*, and *fliJ*

4.2 Development and Performance of Diagnostic Algorithm

The first stage of our QTL algorithm development was to calculate odds ratios across each and every nucleotide position with respect to animal versus host association, and we established background reference sets of 13 animal host bacteria and 15 plant host bacteria respectively that were then aligned. Test cases of 5 animal host bacteria and 3 plant host bacteria were then added to this alignment (Figure 2). The average odds ratios of 10mer regions across the background reference sets and average matching score measures of 10mer regions on the test case sequences are shown in Figure 3. The black line represents the odds of there being a difference between animal and plant host bacteria for the reference sets. These odds values of the black line are going to be added or subtracted to a series of matching scores for the two test sets. Addition happens when the test set matches the animal host bacteria reference set. Subtraction happens when the test set matches the plant host bacteria reference set. The red test case is expected to have more positive scores and the green test case is expected to have more negative scores.

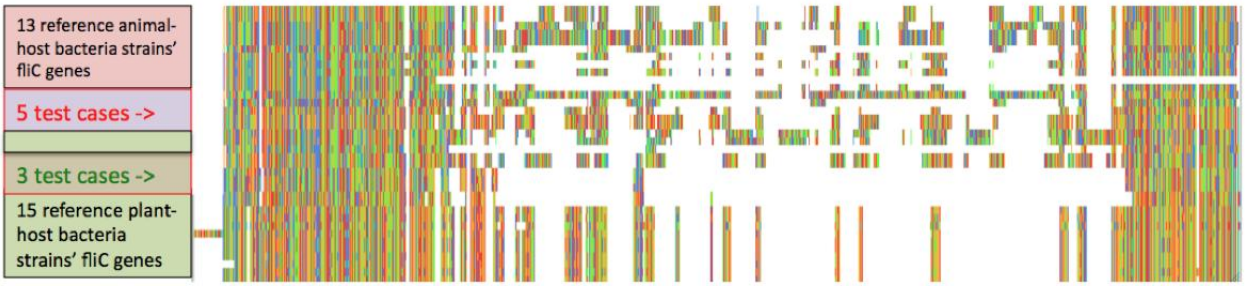

Figure 2. Alignment of background reference sets and test sets for *fliC* gene

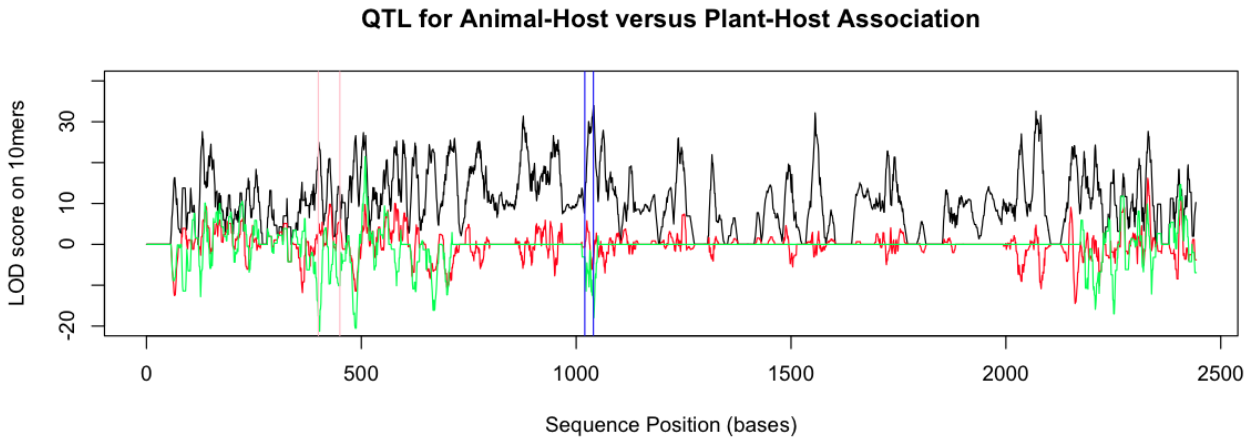

Figure 3. Odds ratios of 10mer regions across the background reference sets (black line), and average matching score measures of 10mer regions on the test case sequences for animal host bacteria (red) and plant host bacteria (green) are shown in Figure 3.

Differentiating regions were identified through visualization of regions where matching score measures were both high for the animal host bacteria test cases and low for plant host bacteria test cases. A leftward differentiating region encoding for an N terminal portion of flagellin was identified at the nucleotide position coordinates of 400-450 (pink lines) and a middle differentiating region was identified at the nucleotide position coordinates 1020-1040 (blue lines).

The 400-450 and 1020-1040 nucleotide coordinate regions from the animal and plant consensus sequences were then identified. The 400-450 nucleotide coordinate regions from the animal and plant consensus sequences were found to be, respectively, "actccattcaggacgaaatcaacccgcgtctggacgaaattgaccgcgtat" and "aagcgctgaactccgaagtcaagcagctcacctcggaaatcgaccgcgtcg." The 1020-1040 nucleotide coordinate regions from the animal and plant consensus sequences were found to be, respectively, "gcagaaaaccagaaagctacg" and "ga--------ccggcatgtac."

These diagnostic consensus sequences were retrospectively tested against the original sets of animal and plant host bacterial strains BLASTN (window size of 7, *E* values < 0.05) for a positive match) for sensitivity, specificity and predictive value (Tables 5 and 6).

Table 7

Performance of animal-host bacteria diagnostic consensus sequences for sensitivity, specificity and predictive value bacterial fliC gene sequences

	Sensitivity (TP/(TP+FN)) Do we get positive matches for all positive (animal host bacteria strain) cases?	**Specificity (TN/(TN+FP))** Do we reject for all negative (plant host bacteria) cases?	**Positive predictive value (TP/(TP+FP))** If a prediction is made, will it be "reliable" (actually true)?
N-terminal associated region pattern (400-450 nt)	0.83	0.11	0.48
Middle region pattern (1020-1040 nt)	0.61	0.83	0.79

Table 8

Performance of plant-host bacteria diagnostic consensus sequences for sensitivity, specificity and predictive value bacterial fliC gene sequences

	Sensitivity (TP/(TP+FN)) Do we get positive matches for all positive (plant host bacteria) cases?	**Specificity (TN/(TN+FP))** Do we reject for all negative (animal host bacteria) cases?	**Positive predictive value (TP/(TP+FP))** If a prediction is made, will it be "reliable" (actually true)?
N-terminal associated region pattern (400-450 nt)	0.77	0.55	0.63
Middle region pattern (1020-1040 nt)	0.61	0.88	0.85

The N-terminal associated region demonstrated greater values of sensitivity (≥0.77) than the middle region (0.61) and the middle region demonstrated greater values of specificity and positive predictive values (≥0.79) than the N-terminal associated region (≤0.55) for retrospective testing against both animal host bacteria and plant host bacteria data sets. (Tables 7 and 8).

4.3 Prospective Study of Agricultural Produce

Lettuce was screened for the presence of the diagnostic consensus sequences to infer for the presence of bacteria with *fliC* genetic markers for animal host association or plant host association. Positive matches (BLASTN, word size = 7, $E<0.05$) for the 400-450 N-terminal genetic marker of plant host-associated bacteria were found in head lettuce from a single grocery store, and this finding was found for two of three replicate samplings from the head lettuce from the single grocery store. This positive match for the same genetic marker was also found for a single replicate of combined sampling from all lettuce specimens put together. An almost positive match (BLASTN, word size = 7, $E=0.054$) was found for the 1020-1040 N-terminal genetic marker of plant host-associated bacteria were found in the same head lettuce specimen from the single grocery store, and for three replicates of combined samplings from all lettuce specimens put together. The other grocery store and other lettuce specimens were negative for any match to the plant and animal 400-450 N-terminal and 1020-1040 N-terminal genetic markers.

Alignments between the targeted Sanger sequencing reactions with the diagnostic consensus sequence were overall very limited. As may have been due to the bulk DNA extraction of both lettuce and microbial DNA, the single-primer sequencing method of the sequencing service, and limitations on the sequencing service's purification protocols, DNA sequencing results were poor. Only patches of high PHRED scores (<20 nucleotides) were observed from the inferred sequencing reads which were generally less than 300 nucleotides in length, and not the expected sequence length of a *fliC* gene (1153 to 1476 nucleotides). The N-terminal associated consensus sequence for plant host bacteria revealed an alignment region of AAGTCAAGCA with the reported sequences, and was found both for the single lettuce

specimen and the combined lettuce specimens. The middle region consensus sequence for plant host bacteria revealed an alignment was for a sequence of GGCATGTAC, and was found both for the single lettuce specimen and the combined lettuce specimens. For a BLAST search of the KEGG database, these sequences identified a joint match with *fliC* for a *Xanthomonas axonopodis, Xanthomonas campestris* or *Xanthomonas citri* bacterial strains.

CHAPTER 5

Discussion and Conclusion

This study represents an effort to develop an analytical approach for matching the gene of a known virulence factor, flagella, to its host association. Potential uses of this analytical approach for both targeted Sanger sequencing and next generation sequencing methods include diagnostics for food safety, investigation of other known virulence factors, and the discovery of genes not yet established to be virulence factors.

With an expanding set of fully sequenced bacterial genomes (Chain et al., 2009; Goetz, 2012), there are both challenges and opportunities for comparing differences in genomic content to phenotype and host-environment associations. A practical challenge is to set up multiple sets of these bacterial genomes with known phenotypic and host-environment association metadata across which genomic content can be analyzed. KEGG and IMG were two tools that, collectively, provided helpful starting points for assembling data (Kanehisa & Goto, 2000; Markowitz et al., 2012). KEGG was used for screening those bacterial strains having all four candidate flagella genes, *fliC*, *fliD*, *fliJ* and *flgG*, as well as retrieving genetic sequence data. IMG provided metadata on host association and pathogenicity across different bacterial strains with fully sequenced genomes that we cross-indexed from KEGG. After verification of bacterial strain host association and pathogenicity based on unambiguous reports in the literature, we established an 18 versus 18 set of bacterial strains associated with animal versus plant hosts respectively. We constrained the comparison to the *Gammaproteobacteria* class (except for *Ralstonia solanacearum* which is from the *Betaproteobacteria* class). There were limited pairwise categorizations across these two sets. The closest taxonomic representation across the two sets was within the common genus of *Pseudomonas.* There was also a broader taxonomic

representation across the two sets for the *Enterobacteriaceae* family (*Erwinia* sp. of plant host association with the animal-host associated *Edwardsiella, Escherichia, Salmonella, Shigella, Yersinia, and Vibrio* sp.). In summary, the establishment of this comparative set must confront cross-indexing and verification across public data repositories, evidence from the literature, and – perhaps in the future – optimize opportunities for pairwise categorizations allowing for closer interpretation and more a powerful statistical design for analysis. This study in the future could be furthered analyzed with increasing data from, for instance, the 100K foodborne pathogen genome project (Goetz, 2012). For studies of the flagellum in contexts other than animal versus plant host association, another interesting comparison would be with nonpathogenic flagellar species such as, for example, *Caulobacter crescentus*. The next steps of the bioinformatics study were to evaluate and compare diversity and conservation across different flagellar genes. We tried different applications for possible discovery workflows such as the clientside GUI-based Jcoda which is fairly integrative for analyzing sequencing conservation (Steinway et al., 2010). Ultimately we settled on step-by-step workflows involving a common core of web-based applications for transforming, visualizing and analyzing the set of sequences flexibly for comparisons of codon-based alignment, dN/dS calculations, primer design, and QTL analysis.

For corresponding *fliC* genetic patterns with animal versus plant host associations, we analyzed variation of the *fliC* gene across 36 strains using a quantitative trait loci approach, and measured sensitivity, specificity and positive predictive values of host-related consensus regions. Based on how different regions demonstrated contrasting degrees of sensitivity and specificity, this suggests a scheme where sensitive regions may be used for initial detection of possible genetic pattern, followed by specific regions that support a more definitive finding. In the study of lettuce, this path of interpretation seemed relevant with the N-terminal associated region

giving the first indication of a plant host bacteria, and the middle region adding to the specificity of the result. Another scenario of flagellar gene sequences and their correspondence with bacterial varieties is horizontal transfer (Schonknecht et al, 2013). We did find some discrepancy between the gene trees and species trees of flagella genes that suggested some level of horizontal transfer could be occurring mainly within the separate groups of animal versus plant host bacteria. The inference of horizontal transfer is not straightforward (Schonknecht et al, 2013), but new enhancements to the phylogeny software application Notung may be a new path for resolving this question across a diverse range of bacteria (Chen, Durand and Farach-Colton, 2000).

We have concluded that out of four genes, *fliC* is the best gene for a diagnostic workflow. *fliC* diversification was most distinctive for animal host-associated bacteria versus plant host-associated bacteria when we compared to the other three genes. From our QTL-based analysis, there were regions of *fliC* that have moderate levels of sensitivity, specificity and positive predictive value – highest for regions within plant host bacteria *fliC* genes. This is perhaps due to the greater conservation of the *fliC* gene across plant host bacteria compared to animal host bacteria. Further developments involving more data and a refined algorithmic approach may enhance how these regions may be used as genetic markers to differentiate for bacteria with different phenotypes of host association. When our results were sent to our sequencing service, we experienced some constraints with the sequencing results. The sequencing service utilized a linear amplification protocol by using one primer (forward primer or reverse primer) instead of exponential amplification protocol, using both primers. In order to increase the predictive value of our approach, we expect that the DNA extraction and targeted PCR steps could be further optimized, or next generation sequencing technologies could be utilized to harvest metagenomic

and whole genome sequencing data sets. Follow-up analyses to identify and validate the specific bacterial strains inferred to be on agricultural and livestock specimens would include further sequence-based analysis and the isolation and culture of bacterial strains. Ultimately, as we may develop this approach to confidently distinguish animal host bacteria from plant host bacteria, this may lead to an innovative, broad-based method that is an essential part of the diagnostic arsenal for food safety.

References

Abby, S. S., & Rocha, E. P. (2012). The non-flagellar type III secretion system evolved from the bacterial flagellum and diversified into host-cell adapted systems. *PLoS Genetics*, *8*(9), e100298.

Alfano, J. R., & Collmer, A. (2004). Type III secretion system effector proteins: double agents in bacterial disease and plant defense. *Annu. Rev. Phytopathol.*, *(42*)385-414.

Arora, S. K., Ritchings, B. W., Almira, E. C., Lory, S., & Ramphal, R. (1998). The *Pseudomonas aeruginosa* flagellar cap protein, FliD, is responsible for mucin adhesion. *Infection and immunity*, *66*(3), 1000-1007.

Boch, J., & Bonas, U. (2001). Gram negative plant pathogenic bacteria. *Emerging bacterial pathogens*, Vol (8), 186-196.

Brooke, J. S. (2012). *Stenotrophomonas maltophilia*: an emerging global opportunistic pathogen. *Clinical microbiology reviews*, *25*(1), 2-41.

Camper, A. K., Hayes, J. T., Sturman, P. J., Jones, W. L., & Cunningham, A. B. (1993). Effects of motility and adsorption rate coefficient on transport of bacteria through saturated porous media. *Applied and environmental microbiology*, *59*(10), 3455-3462.

Cavalier-Smith, T. (2002). The neomuran origin of archaebacteria, the negibacterial root of the universal tree and bacterial megaclassification. *International Journal of Systematic and Evolutionary Microbiology*, *52*(1), 7-76.

Chain, P. S. G., Grafham, D. V., Fulton, R. S., Fitzgerald, M. G., Hostetler, J., Muzny, D., ... & Detter, J. C. (2009). Genome project standards in a new era of sequencing. *Science (New York, NY)*, *326*(5950).

Chen, K., Durand, D., & Farach-Colton, M. (2000). NOTUNG: a program for dating gene duplications and optimizing gene family trees. *Journal of Computational Biology*, *7*(3-4), 429-447.

Costerton, J. W., Stewart, P. S., & Greenberg, E. P. (1999). Bacterial biofilms: a common cause of persistent infections. *Science*, *284*(5418), 1318-1322.

Duncan, C. G., Leary, R. J., Lin, J. C., Cummins, J., Di, C., Schaefer, C. F., ... & Yan, H. (2009). Identification of microbial DNA in human cancer. *BMC medical genomics*, *2*(1), 22.

Felix, G., Duran, J.D., Volko, S., & Boller, T. (1999). Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *The Plant Journal*, *18*(3), 265-276.

Fraser, G. M., González-Pedrajo, B., Tame, J. R., & Macnab, R. M. (2003). Interactions of FliJ with the *Salmonella* type III flagellar export apparatus. *Journal of bacteriology*, *185*(18), 5546-5554.

Gerlach, R. G., & Hensel, M. (2007). Protein secretion systems and adhesins: the molecular armory of Gram-negative pathogens. *International Journal of Medical Microbiology*, *297*(6), 401-415.

Goetz, G. (2012). Genome Sequencing of 100,000 Foodborne Pathogens Underway. *Food Safety News*. July 13, 2012 (http://www.foodsafetynews.com/2012/07/genome-sequencing-of-100000-foodborne-pathogens-underway; accessed 2014/06/06)

González-Pedrajo, B., Fraser, G. M., Minamino, T., & Macnab, R. M. (2002). Molecular dissection of *Salmonella* FliH, a regulator of the ATPase FliI and the type III flagellar protein export pathway. *Molecular microbiology*, *45*(4), 967-982.

Gupta, R. S. (2011). Origin of diderm (Gram-negative) bacteria: antibiotic selection pressure rather than endosymbiosis likely led to the evolution of bacterial cells with two membranes. *Antonie Van Leeuwenhoek*, *100*(2), 171-182.

Gurlitz, R. H., Lamb, P. W., & Matthysse, A. G. (1987). Involvement of carrot cell surface proteins in attachment of *Agrobacterium tumefaciens*. *Plant physiology*, *83*(3), 564-568.

Haley, B. J., Choi, S. Y., Hasan, N. A., Abdullah, A. S. H., Cebula, T. A., Huq, A., & Colwell, R. R. (2013). Genome sequences of clinical *Vibrio cholerae* isolates from an oyster-borne *cholera* outbreak in Florida. *Genome Announcements*, *1*(6), e00966-13.

Hammond-Kosack, K. E., & Jones, J. D. (1997). Plant disease resistance genes. *Annual Review of Plant Biology*, *48*(1), 575-607.

Homma, M., DeRosier, D. J., & Macnab, R. M. (1990). Flagellar hook and hook-associated proteins of *Salmonella typhimurium* and their relationship to other axial components of the flagellum. *Journal of molecular biology*, *213*(4), 819-832.

Homma, M., Kutsukake, K., Hasebe, M., Iino, T., & Macnab, R. M. (1990). FlgB, FlgC, FlgF and FlgG: a family of structurally related proteins in the flagellar basal body of *Salmonella typhimurium*. *Journal of Molecular Biology*, *211*(2), 465-477.

Hunt, K. M., Foster, J. A., Forney, L. J., Schütte, U. M., Beck, D. L., Abdo, Z., ... & McGuire, M. A. (2011). Characterization of the diversity and temporal stability of bacterial communities in human milk. *PLoS One*, *6*(6), e21313.

Ibuki, T., Imada, K., Minamino, T., Kato, T., Miyata, T., & Namba, K. (2011). Common architecture of the flagellar type III protein export apparatus and F-and V-type ATPases. *Nature structural & molecular biology*, *18*(3), 277-282.

Jenneman, G. E., McInerney, M. J., & Knapp, R. M. (1985). Microbial penetration through nutrient-saturated Berea sandstone. *Applied and environmental microbiology*, *50*(2), 383-391.

Jin, Q., Hu., Brown, I., McGhee, G., Hart, P., Jones, A.L., & He, S.Y. (2001). Visualization of secreted Hrp and Avr proteins along the Hrp pilus during type III secretion in *Erwinia amylovora* and *Pseudomonas syringae. Molecular Microbiology, 40*(5), 1129-1139.

Joys, T. M. (1985). The covalent structure of the phase-1 flagellar filament protein of *Salmonella typhimurium* and its comparison with other flagellins. *Journal of Biological Chemistry*, *260*(29), 15758-15761.

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, *28*(1), 27-30.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, *30*(4), 772-780.

Kearsey, M. J. (1998). The principles of QTL analysis (a minimal mathematics approach). *Journal of Experimental Botany*, *49*(327), 1619-1623.

Kim, M. C. (1989). Molecular characterization and regulation of the gene expression of offlhB operon in *Eschericha coli*. (Order No. 9015751, University of Illinois at Chicago). *ProQuest Dissertations and Theses*, , 210-210 p.

Khan, S., & Macnab, R. M. (1980). Proton chemical potential, proton electrical potential and bacterial motility. *Journal of molecular biology*, *138*(3), 599-614.

Komeda, Y. (1982). Fusions of flagellar operons to lactose genes on a mu lac bacteriophage. *Journal of bacteriology*, *150*(1), 16-26.

Korber, D. R., Lawrence, J. R., & Caldwell, D. E. (1994). Effect of motility on surface colonization and reproductive success of *Pseudomonas fluorescens* in dual-dilution continuous culture and batch culture systems. *Applied and environmental microbiology*, *60*(5), 1421-1429.

Kutsukake, K., Ohya, Y., & Iino, T. (1990). Transcriptional analysis of the flagellar regulon of *Salmonella typhimurium. Journal of bacteriology*, *172*(2), 741-747.

Lawrence, J. R., Delaquis, P. J., Korber, D. R., & Caldwell, D. E. (1987). Behavior of *Pseudomonas fluorescens* within the hydrodynamic boundary layers of surface microenvironments. *Microbial ecology*, *14*(1), 1-14.

Lillehoj, E. P., Kim, B. T., & Kim, K. C. (2002). Identification of *Pseudomonas aeruginosa* flagellin as an adhesin for Muc1 mucin. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, *282*(4), L751-L756.

Lynch, M. (2012). The evolution of multimeric protein assemblages. *Molecular biology and evolution*, *29*(5), 1353-1366.

Macnab, R. M. (1992). Genetics and biogenesis of bacterial flagella. *Annual review of genetics*, *26*(1), 131-158.

Macnab, R. M. (2004). Type III flagellar protein export and flagellar assembly. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, *1694*(1), 207-217.

Mahenthiralingam, E., & Speert, D. P. (1995). Nonopsonic phagocytosis of *Pseudomonas aeruginosa* by macrophages and polymorphonuclear leukocytes requires the presence of the bacterial flagellum. *Infection and immunity*, 63(11), 4519-4523.

Markowitz, V. M., Chen, I. M. A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., ... & Kyrpides, N. C. (2012). IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic acids research*, *40*(D1), D115-D122.

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y.M., Buso, N., Cowley, A.P., & Lopez, R. (2013) Analysis Tool Web Services from the EMBL-EBI. *Nucleic acids research 2013 Jul;41(Web Server issue):W597-600* doi:10.1093/nar/gkt376

Minamino, T., Chu, R., Yamaguchi, S., & Macnab, R. M. (2000). Role of FliJ in Flagellar Protein Export in *Salmonella*. *Journal of Bacteriology*, *182*(15), 4207-4215.

Minamino, T., & Macnab, R. M. (1999). Components of the *Salmonella* flagellar export apparatus and classification of export substrates. *Journal of Bacteriology, 181(*5), 1388-1394.

Mueller, R. F. (1996). Bacterial transport and colonization in low nutrient environments. *Water Research*, *30*(11), 2681-2690.

Mueller, R. F., Characklis, W. G., Jones, W. L., & Sears, J. T. (1992). Characterization of initial events in bacterial surface colonization by two *Pseudomonas* species using image analysis. *Biotechnology and Bioengineering*, *39*(11), 1161-1170.

Newell, D. G., McBride, H., & Dolby, J. M. (1985). Investigations on the role of flagella in the colonization of infant mice with *Campylobacter jejuni* and attachment of *Campylobacter jejuni* to human epithelial cell lines. *Epidemiology and Infection*, *95*(2), 217-227.

O'Toole, G., & Kaplan, H. B. (2000). Biofilm Formation as Microbial Development. *Annual Review of Microbiology*, *54*(1), 49.

Palmer, K. L., Aye, L. M., & Whiteley, M. (2007). Nutritional cues control *Pseudomonas aeruginosa* multicellular behavior in cystic fibrosis sputum. *Journal of bacteriology*, *189*(22), 8079-8087.

Poggio, S., Abreu-Goodger, C., Fabela, S., Osorio, A., Dreyfus, G., Vinuesa, P., & Camarena, L. (2007). A complete set of flagellar genes acquired by horizontal transfer coexists with the endogenous flagellar system in *Rhodobacter sphaeroides*. *Journal of bacteriology*, *189*(8), 3208-3216.

Proft, T., & Baker, E. N. (2009). Pili in Gram-negative and Gram-positive bacteria—structure, assembly and their role in disease. *Cellular and molecular life sciences*, *66*(4), 613-635

Rescigno, M., Citterio, S., Thèry, C., Rittig, M., Medaglini, D., Pozzi, G., ... & Ricciardi-Castagnoli, P. (1998). Bacteria-induced neo-biosynthesis, stabilization, and surface expression of functional class I molecules in mouse dendritic cells. *Proceedings of the National Academy of Sciences*, *95*(9), 5229-5234.

Reynolds, P. J., Sharma, P., Jenneman, G. E., & McInerney, M. J. (1989). Mechanisms of microbial movement in subsurface materials. *Applied and environmental microbiology*, *55*(9), 2280-2286.

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics 16*, (6) pp276—277

Romantschuk, M. (1992). Attachment of plant pathogenic bacteria to plant surfaces. *Annual review of phytopathology*, *30*(1), 225-243.

Schirmer, T. (1998). General and specific porins from bacterial outer membranes. *Journal of structural biology*, *121*(2), 101-109.

Smith, E. E., Buckley, D. G., Wu, Z., Saenphimmachak, C., Hoffman, L. R., D'Argenio, D. A., ... & Olson, M. V. (2006). Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. *Proceedings of the National Academy of Sciences*, *103*(22), 8487-8492.

Steinway, S. N., Dannenfelser, R., Laucius, C. D., Hayes, J. E., & Nayak, S. (2010). JCoDA: a tool for detecting evolutionary selection. *BMC bioinformatics*,*11*(1), 284.

Mikita, S., David, T., & Peer, B. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res. 34*, W609-W612.

Taylor, J. D., Teverson, D. M., Allen, D. J., & Pastor-Corrales, M. A. (1996). Identification and origin of races of *Pseudomonas syringae* pv. *phaseolicola* from Africa and other bean growing areas. *Plant Pathology*, *45*(3), 469-478.

Tasteyre, A., T. Karjalainen, V. Avesani, M. Delmee, A. Collignon, P. Bourlioux, & M. C. Barc. (2000). Phenotypic and genotypic diversity of the flagellin gene (*fliC*) among *Clostridium difficile* isolates from different sero- groups. *J. Clin. Microbiol. 38*:3179–3186.

Tasteyre, A., Barc, M. C., Collignon, A., Boureau, H., & Karjalainen, T. (2001). Role of FliC and FliD flagellar proteins of *Clostridium difficile* in adherence and gut colonization. *Infection and Immunity*, *69*(12), 7937-7940.

Tasteyre, A., Karjalainen, T., Avesani, V., Delmée, M., Collignon, A., Bourlioux, P., & Barc, M. C. (2001). Molecular characterization of *fliD* gene encoding flagellar cap and its expression among *Clostridium difficile* isolates from different serogroups. *Proceedings of the National Academy of Sciences*, *39*(3), 1178-1183.

Turnbull, G. A., Morgan, J. A. W., Whipps, J. M., & Saunders, J. R. (2001). The role of motility in the in vitro attachment of *Pseudomonas putida* PaW8 to wheat roots. *FEMS microbiology ecology*, *35*(1), 57-65.

Van Delden, C., & Iglewski, B. H. (1998). Cell-to-cell signaling and *Pseudomonas aeruginosa* infections. *Emerging infectious diseases*, *4*(4), 551.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A, Clamp, M. & Barton, G. J. (2009). Jalview Version 2 - a multiple sequence alignment editor and analysis. *Workbench Bioinformatics*, *25*(9) 1189-1191 doi: 10.1093/bioinformatics/btp033

Delport, W., Poon, A. F., Frost, S. D., & Pond, S. L. K. (2010). Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics*, *26*(19), 2455-2457.

Willis, D. K., Hrabak, E. M., Rich, J. J., Barta, T. M., Lindow, S. E., & Panopoulos, N. J. (1990). Isolation and characterization of a *Pseudomonas syringae* pv. *syringae* mutant deficient in lesion formation on bean. *Molecular plant-microbe interactions*, *3*(3), 149-156.

Wilson, D. R., & Beveridge, T. J. (1993). Bacterial flagellar filaments and their component flagellins. *Canadian journal of microbiology*, *39*(5), 451-472.

Wilson, M., Campbell, H. L., Ji, P., Jones, J. B., & Cuppels, D. A. (2002). Biological control of bacterial speck of tomato under field conditions at several locations in North America. *Phytopathology*, *92*(12), 1284-1292.

Winn, W. C., Koneman, C., & Elmer, W. (2006). Koneman's color atlas and textbook of diagnostic microbiology. Philadelphia : Lippincott Williams & Wilkins.

Winsor, G. L., Lam, D. K., Fleming, L., Lo, R., Whiteside, M. D., Nancy, Y. Y., ... & Brinkman, F. S. (2011). *Pseudomonas* Genome Database: improved comparative analysis and

population genomics capability for *Pseudomonas* genomes. *Nucleic acids research*, *39*(suppl 1), D596-D600.

Witt, M. E., Dybas, M. J., Worden, R. M., & Criddle, C. S. (1999). Motility-enhanced bioremediation of carbon tetrachloride-contaminated aquifer sediments. *Environmental science & technology*, *33*(17), 2958-2964.

Worlitzsch, D., Tarran, R., Ulrich, M., Schwab, U., Cekici, A., Meyer, K. C., ... & Döring, G. (2002). Effects of reduced mucus oxygen concentration in airway *Pseudomonas* infections of cystic fibrosis patients. *Journal of Clinical Investigation*, *109*(3), 317-325.

Yancey, R. J., Willis, D. L., & Berry, L. J. (1979). Flagella-induced immunity against experimental *cholera* in adult rabbits. *Infection and immunity*, *25*(1), 220-228.

Yokoseki, T., Kutsukake, K., & Ohnishi, K. (1995). Functional analysis of the flagellar genes in the fliD operon of *Salmonella typhimurium*. *Microbiology*, *141*(7), 1715-1722.

Shute, N. (2013). Your kitchen spices can often harbor *Salmonella* . *National Public Radio*. Retrieved from http://www.npr.org/blogs/thesalt/2013/08/28/216550862/your-kitchen-spices-can-often-harbor-salmonella

Appendix

A pilot study was conducted of comparative host cell attachment of *Pseudomonas aeruginosa* PA01 and *Pseudomonas syringae* DC3000 to lettuce tissue (1 g, manually ground with fingers) and 100 uL directly from culture of C1C17 mouse cells (Figures A1-A4). Qualitative analysis was based on visual assessment of growth density and green pigmentation (potentially pyocyanin – observed in both mouse and lettuce cell treatments) from *Pseudomonas syringae* and *Pseudomonas aeruginosa*. In all instances, bacteria were found in both the unattached and potentially attached layers of the experiment. There were higher levels of green pigmentation that we observed for *P. syringae* attached to lettuce tissue than for *P. aeruginosa* (Figure A4). For one of the petri dishes, bacteria that were in the unattached layer of the experiment with mouse cell lines were found to be more dense for *P. aeruginosa* than *P. syringae* (Figure A1), which may be notable due to the greater CFU density of *P. syringae* used in inoculation. Overall, these preliminary results suggest that separations of a liquid phase may yield levels of bacteria in layers of the liquid with and without eukaryotic cells, but additional steps were not performed to verify actual attachment.

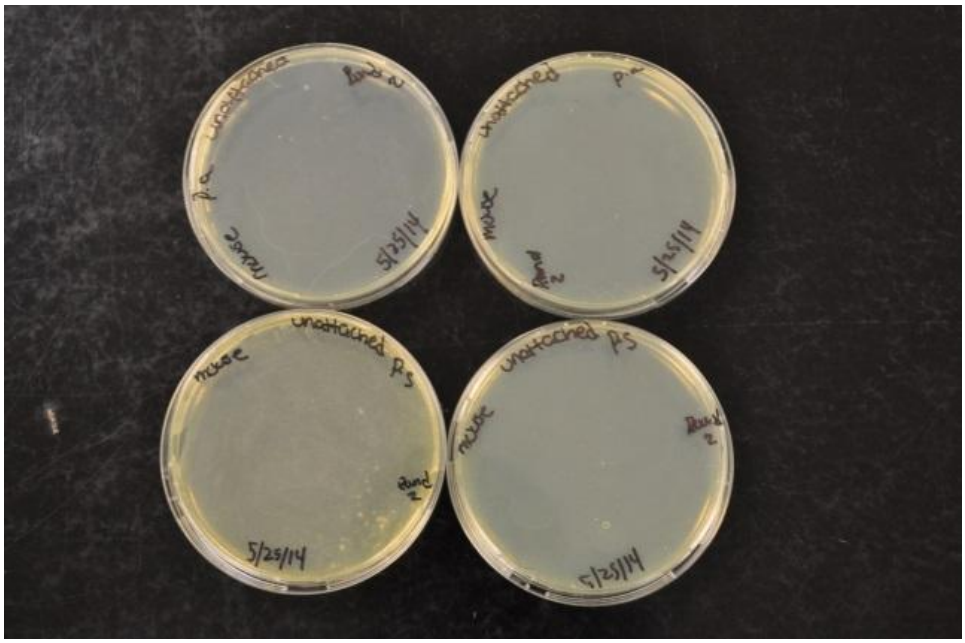

Figure A1. Plating of bacteria that were unattached to mouse cells. Top: *Pseudomonas aeruginosa* PA01; Bottom: *Pseudomonas syringae* DC3000.

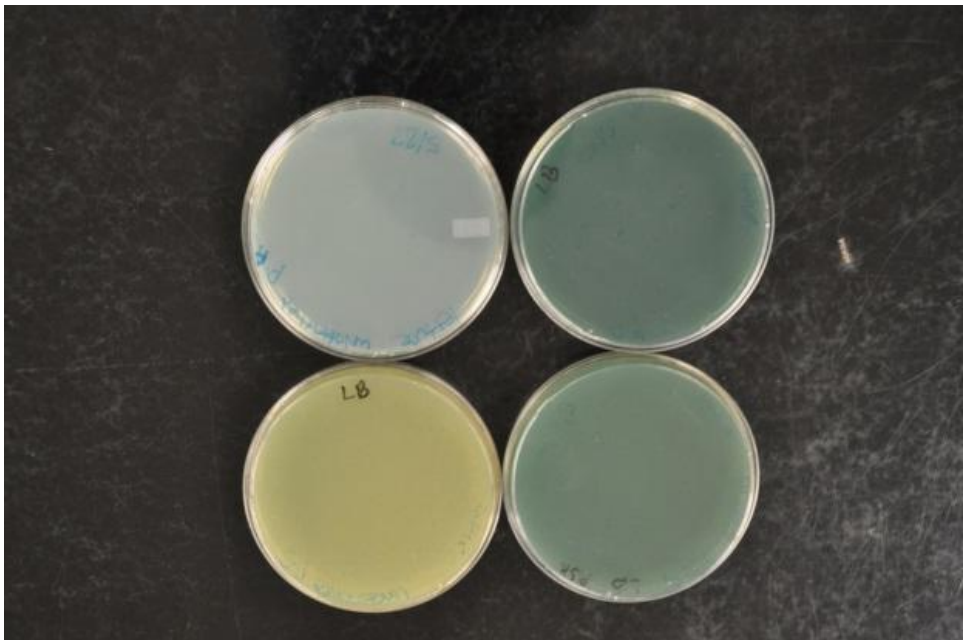

Figure A2. Plating of bacteria that were unattached to lettuce cells. Top: *Pseudomonas aeruginosa* PA01; Bottom: *Pseudomonas syringae* DC3000.

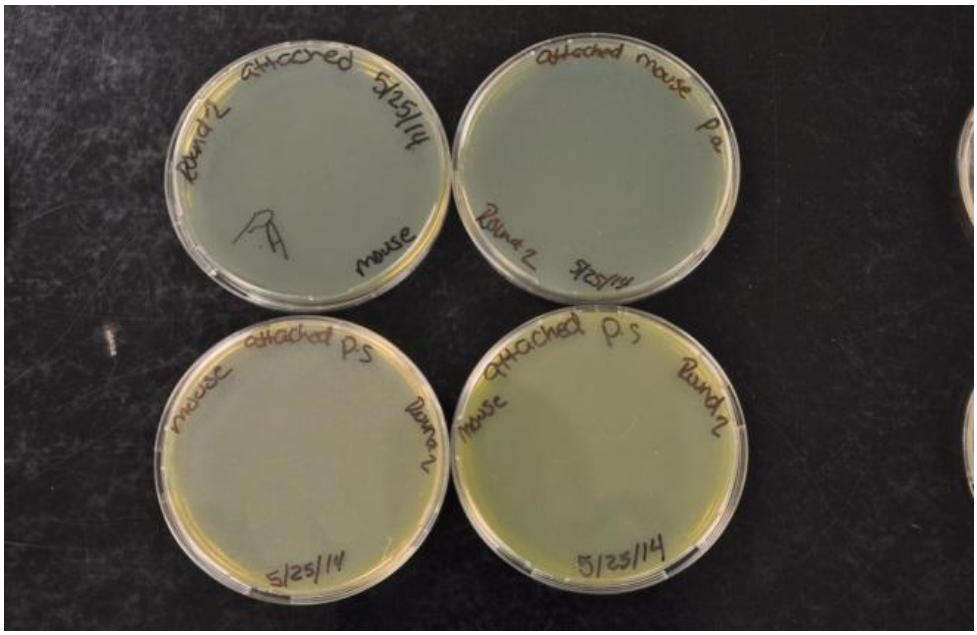

Figure A3. Plating of bacteria that were potentially attached to mouse cells. Top: *Pseudomonas aeruginosa* PA01; Bottom: *Pseudomonas syringae* DC3000.

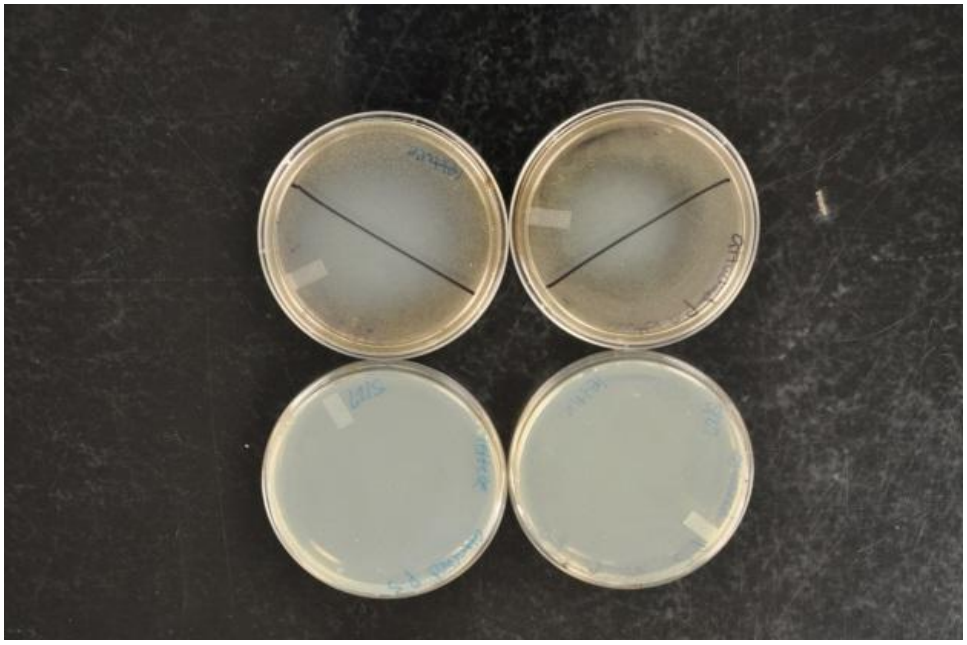

Figure A4. Plating of bacteria that were potentially attached to lettuce cells. Top: *Pseudomonas aeruginosa* PA01; Bottom: *Pseudomonas syringae* DC3000.